Introduction

- Smooth convex minimization: \( \min_x f(x) \)
- \( f \) is convex, \( \nabla f \) is Lipschitz, and \( \mathcal{X} \) is compact.
- Continuous-time dynamics.

Contributions

- Analysis of stochastic dynamics
- Different regimes: persistent noise and vanishing noise
- Acceleration: faster convergence in vanishing noise regime

Stochastic Accelerated Mirror descent

Mirror descent [1]

\[
\begin{align*}
\frac{d x(t)}{dt} &= -\alpha(t) \nabla f(x(t)) \\
\frac{d z(t)}{dt} &= \nabla \psi(z(t))
\end{align*}
\]

Lipschitz mirror map \( \nabla \psi \) maps dual space to \( \mathcal{X} \)

Lyapunov functions: Bregman divergence

\[
D_\psi(x^*, z(t)) = D_\psi(z(t), z^*)
\]

Accelerated Mirror descent [2]

\[
\begin{align*}
\frac{d x(t)}{dt} &= -\alpha(t) \nabla f(x(t)) \\
\frac{d z(t)}{dt} &= \frac{1}{\tau} \int_0^t w(z(\tau)) d\tau
\end{align*}
\]

Averaging in the primal space

Special case: Nesterov's ODE [3] \( \leftarrow \)

Stochastic Accelerated Mirror descent

\[
\begin{align*}
\frac{d z(t)}{dt} &= -\alpha(t) \nabla f(X(t)) = \sigma(X(t), t) dB(t) \\
X(t) &= \frac{1}{\tau} \int_0^t w(z(\tau)) d\tau
\end{align*}
\]

- \( B(t) \) Brownian motion
- \( \sigma(x, t) \) volatility matrix, with \( \text{sup}_{x \in \mathcal{X}} \| \sigma(x, t) \rho(x, t)^T \|_2 \leq \sigma^2(t) \)

Choosing the weights in the vanishing noise regime

Given \( \tau_\epsilon(t) \rightarrow \epsilon \)

choose \( \alpha(t) \rightarrow \epsilon \), \( \beta \rightarrow 1/\epsilon \)

Resulting rate

\[
E[f(X(t))] - f(x^*) \leq O(\epsilon^{-\frac{1}{2}})
\]

Averaging weights

Accelerated dynamics may diverge in the presence of noise, without proper weight rectification.

Analysis and convergence rates

Deterministic \( \tau_\epsilon(t) = \epsilon \)

Stochastic

\[
E(t) = \int_0^t \frac{1}{\epsilon^n} \| f(x(t)) - f(x^*) \|_* + D_\psi(z(t), z^*)
\]

By Itô's formula

\[
dE \leq -\frac{1}{2} \text{tr}(\sigma^2(t) \psi''(Z(t)) \nu(t)) dt
\]

We also show a.s. asymptotic rate

\[
f(X(t)) - f(x^*) \leq \mathcal{O} \left( \frac{\text{log}(1/k)}{\sqrt{\text{log}(1/k)}} \right)
\]

where \( k \rightarrow \epsilon \)

Contributions

- Analysis of stochastic dynamics
- Different regimes: persistent noise and vanishing noise
- Acceleration: faster convergence in vanishing noise regime

Additional results, open questions

Results

- A.S. convergence if \( \alpha(t) \sqrt{\log t} \rightarrow 0 \) and \( \int_0^t \alpha^2(t) dt \rightarrow \nu \)

Open questions

- Discretization
- Faster rates for strongly convex functions
- Multiplicative models of noise: \( \text{sup} \| \sigma(x) \sigma(x)^T \|_2 \), scales with \( \| \nabla f(x) \|_2^2 \)